

AD-A275 489



①

Control Question Test
1

DTIC
ELECTE
FEB 02 1994
S A

Control Question Tests by Police and Laboratory
Polygraph Operators on a Mock Crime and Real Events.

M. T. Bradley, M. C. Cullen,

S. B. Carle

University of New Brunswick Saint John

This document has been approved
for public release and sale; its
distribution is unlimited.

Running head: POLICE AND LABORATORY POLYGRAPHERS COMPARED

94 2 01 15 I

748

94-03335



**Best
Available
Copy**

Abstract

Males and females, truthful or deceptive, about a real life embarrassing story or a laboratory mock crime were examined with Control Question detection of deception tests. Exams were conducted either by a police or a laboratory trained polygraph operator. Subjects were more reactive to event relevant questions when deceptive than when truthful. Police scored subject records more towards innocence whereas laboratory investigators scored them more towards guilt. This was especially pronounced with SRR measurement on embarrassing stories. Such a result could mean that laboratory investigators when mistaken would have a tendency to classify innocent people as guilty when dealing with real events whereas the police when wrong would tend to classify the guilty as innocent.

DTIC QUALITY INSPECTED 2

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Code	
Dist	Avail and/or Special
A-1	

Control question tests by police and laboratory polygraph operators on a mock crime and real events.

A number of attempts to assess criterion validity of the Control Question Test have been made in both field and laboratory studies. A recent review (Ben-Shakhar & Furedy, 1990), suggested that validity issues have not been fully resolved because of various problems particular to each area of study. A major problem in the field is that it is difficult to verify who is actually guilty or innocent by any satisfactory criteria outside of the polygraph test. Therefore, test accuracy levels cannot be determined with confidence.

This particular problem is avoided in laboratory studies because subjects can be assigned to their conditions but other problems arise. Laboratory studies are simulations of crimes. Usually, these simulations involve relatively small incentives and the population (students) participates in an exercise that may not generalize to field situations (Saxe,

Dougherty, & Cross, 1985). Laboratory examiners are typically highly trained in psychology, psychophysiological measurement and general testing, whereas field investigators are specifically trained for criminal polygraph work and have general criminal interrogation experience.

Bradley and Cullen (1993) selected one area of difference and attempted to add realism to the laboratory situation by examining events that had actually occurred to subjects. Subjects were asked to provide, from their own life, an embarrassing story that had a strong emotional impact on them. The story, which for ethical reasons, had to be non-criminal, involved events that subjects preferred no-one knew of and they would rather deny. Subjects, examined with the Control Question Test on two stories, one in which they were the principal actor and one in which they had no part, were accurately classified as deceptive in denying their own story and as truthful when denying another story.

The present study furthered explored the use of real events by comparing the results of real event

examinations with those from a mock crime situation. In addition, both police and laboratory trained examiners tested subjects.

The two police officers had been trained by the Ottawa Canadian Police College in the early 1980's. Their work since that training has been in the use of the CQT for criminal investigation. In a comparable way to the laboratory examiners, the police officers agreed to blindly examine subjects solely on the basis of knowing only the details of the mock crime or the particular embarrassing story to classify whether subjects were deceptive or truthful about their role in these events. Beyond that, the police were free to apply the CQT in the way that their training and experience dictated that they should. The laboratory examiners were restricted to a laid out protocol.

The scores of subjects examined on embarrassing stories were compared with those of subjects examined in a typical mock crime situation. This provided a direct test of conditions hopefully closer to actual field conditions as versus the enacted artificial mock crime. If considerations by Iacono and Patrick (1988)

are correct then, the accuracy of detection rates from embarrassing incidents should be somewhat less than those found in the mock crime situation.

To find out if training or experience makes a difference, results from subjects examined by police polygraph operators were compared with those tested by laboratory trained operators.

Method

Subjects

One hundred and twenty male and female introductory psychology student volunteers participated for a bonus course credit. Prior to volunteering, they were, through a consent sheet, informed of most of the experimental requirements. A sensitive issue highlighted in the form involved the fact that a limited number of people who assisted with the experiment would be able to associate their name with their embarrassing incident.

Apparatus

A Lafayette model 760-566 polygraph was used to record skin resistance responses (SRR) and respiration. Skin resistance was measured by standard Lafayette zinc chloride electrodes. After the skin had been cleaned with a cotton swab dipped in alcohol, the electrodes were attached to the medial phalanges of the first and second fingers. Respiration, both thoracic and abdominal, were measured by a standard Lafayette pneumatic chest assembly. Baseline and sensitivity levels were adjusted individually.

Procedure

Forty three male and female volunteer subjects were asked to write out in some detail a very embarrassing incident in which they were involved. The stories were read for clarity and understanding. The authors of the thirty selected stories were contacted and polygraph examination sessions were arranged. An equal number of subjects were contacted who had not written a story. They were examined on one of stories generated by the first group of subjects. Subjects who appeared truthful on the subsequent polygraph test

received \$20.00.

A second set of subjects followed instructions leading them to be guilty or innocent of a mock crime murder. Guilty subjects were asked to go into a room labelled hotel, pick up a gun from the window ledge, and shoot a mannequin wearing a red shirt three times in the chest. The mannequin was wearing a name tag with "Bob" written on it and had \$15 in his shirt pocket. Guilty subjects stole the \$15, put the money in their footwear, hid the gun in a wastebasket and left the room. They had about 10 minutes to complete their instructions and once done they went go to a room to await the return of an experimenter who arranged for a polygraph test. Subjects were told that if they appeared innocent of the crime they would receive \$20.00.

The instructions for the innocent subjects informed them that they were murder suspects and, although they had no alibi to account for their activities, they were given a chance to prove their innocence on the polygraph. These subjects were informed that they would receive \$20.00 for a judgment of innocent on the

polygraph test.

All subjects were reminded that during the polygraph examination they were, depending on the group, to deny their involvement in the mock crime or the embarrassing incident. In that way half of the subjects were deceptive and half were truthful about the events. Subjects were reminded at this point that they could receive \$20.00 for a judgment of innocent whereas they would not receive money if they were judged guilty.

The polygraph examinations, done by a laboratory examiner blind to the whether subjects were attempting deception or not, were standard for everyone in the mock crime groups. The examinations were individually constructed for each subject in the embarrassing event groups. For example, a subject was examined on vomiting over 1st class passengers in an airplane while he was running for the washroom. The Control Question Test was as follows:

- (1) Your first name is _____?
- (2) Are you going to tell me the entire truth in this examination?
- (3) Do you believe I will only ask the questions

that we have reviewed in this test?

- (4) Have you ever been arrested?
- (5) Did your embarrassing event take place on an airplane?
- (6) Have you ever cheated on your income tax?
- (7) Were you walking through the first class compartment when you vomited?
- (8) Is your last name _____?
- (9) Did you ever wish to seriously hurt a friend?
- (10) Did you make a mess on the passengers in the plane?

The standard mock crime Control Question Test was as follows:

- 1. Is your last name _____?
- 2. Regarding the crime , do you intend to answer all questions about it truthfully ?
- 3. Are you convinced that I will only ask questions on this test that we have already okayed ?
- 4. Before the age of 18, did you ever steal any money ?
- 5. Did you take the money from the body of the man?

6. While you were in school, did you ever steal anything of value ?
7. While you were in the hotel, did you shoot the man?
8. Is your first name _____?
9. Before the age of 18, did you ever physically harm anyone in any way ?
10. Did you use the gun to kill the man ?

The initial three questions and question number 8 pertained to general issues. Questions 4, 6, and 9 were control questions and questions 5, 7, and 10 were event relevant. Control questions focus on possibly incriminating issues that are not the true concern of the investigator. They are meant to be emotionally evocative, however, because they are about issues that, in this study, are related to other embarrassing situations.

The question set during the actual examination was repeated three times. After each question, approximately 20 seconds was allowed for physiological responses.

The police officers did not follow their normal

procedure involving pretest interviews as they had no investigative evidence about the suspect. They modified the mock crime CQT by having all three crime relevant questions concentrate on what they considered the single salient issue, the taking of the money. To illustrate, the following are the crime relevant questions from one test. "Concerning the case, did you take the money belonging to Bob?;, do you have Bob's money in your shoe?;, are you hiding Bob's money in your shoe?".

In a similar fashion, in general, the police focused on a single issue with embarrassing stories. The following example illustrates crime relevant questions from a story. "Regarding that story, did your mom use your condoms to embarrass you in front of your friends?;, was it your mom who embarrassed you by blowing up your hidden condoms?,, did you get embarrassed when your mom blew-up your hidden condoms?"

Data Analysis

The major analyses involved 2 x 2 x 2 MANOVAs and univariate analyses on detection scores derived from

the polygraph recordings of abdominal and thoracic respiration and skin resistance. Gender, situations (mock crime, embarrassing events), and condition (innocent or guilty) were the factors analyzed. Significance for all analyses was accepted at the .05 level.

Skin resistance responses were measured at the maximum decrease in resistance in millimetres occurring within 10 seconds of the beginning of the question. To derive a numerical score responses for control and event/mock crime relevant questions were considered in pairs; the pairs being questions 4 and 5, 6 and 7, and 9 and 10. Depending on whether the size of a response to a control question was two, three, or up to four times larger than the response to the paired event-related question, a positive one, two or three was assigned to the pair. If the response to the event related question was larger, then, depending on the relative difference a negative one, two, or three was assigned to the pair. An alternate method reported in the classification table under SRR1 involved ignoring the magnitude of the difference and the simple

assignment of a +/-1 if there was a difference.

Respiration scores were derived through the use of a contour map wheel. The wheel was used to follow the curvilinear tracings that represented inhalation and exhalation and gave distance readings in millimetres. Measures were taken for 10s of chart time following question onset. Timm (1982) found respiratory suppression associated with deception. If the response to a control question was shorter than to its paired event/mock crime relevant question a +1 was assigned, if longer then a -1 assigned, and if there was no difference the score was 0.

With three sets of questions repeated three times, for each of the measures a total of 9 judgments were made and the scores had the possibility of ranging from +9 (the maximum innocent score) to -9 (the maximum guilt score). If subjects scored greater than +2 they were classed as innocent; less than -2 resulted in a guilty classification. Scores between these numbers were judged as inconclusive. When a composite of measures was created by the police or by laboratory examiners +/-6 was used for the cut points.

Results

Four factor analyses of variance were used to examine three dependent measures. The factors were examiners (police or lab), gender, situation (mock crime or embarrassing story) and condition (guilt or innocence). The dependent measures were SRR scores, thoracic respiration scores and a composite of scores. Because of differences in scoring techniques and measures the composite score for the police consisted of the sum of SRR scores plus a blended thoracic and abdominal score plus a score derived from blood pressure measurements. The composite in the laboratory involved SRR scores, and separate scores from both thoracic and abdominal respiration.

With composite scores, there was an examiners effect ($F(1, 104) = 5.54$ such that subjects tested by police obtained scores more in the positive direction ($M = 2.60$) than subjects tested by laboratory examiners ($M = -.71$). Condition effects showed that the scores of guilty subjects ($M = -4.78$) were more negative than those of innocent subjects ($M = 5.57$), ($F(1, 104) =$

60.81). No other main effects or interactions were significant.

SRR scores differed depending on whether the police ($M = 1.98$) or laboratory personnel ($M = -.40$) conducted the tests, $F(1, 104) = 7.22$. Mock crime subjects were scored in a more positive direction ($M = 1.35$) than embarrassing event subjects ($M = -.57$). Guilty subjects scored in the negative direction ($M = -2.33$) whereas innocent subjects scored in the positive direction ($M = 3.12$), $F(1, 104) = 42.74$. Embarrassing stories and mock crimes interacted with who conducted the exam $F(1, 104) = 4.38$ (see figure 1).

Figure 1 about here

Simple main effects analysis showed that laboratory examiners scored embarrassing story subjects more negatively than they scored mock crime subjects or than the police scored either type of scenario.

Respiration scores (combined by the police) differed between guilty subjects ($M = -1.30$) and innocent subjects ($M = .75$), $F(1, 104) = 16.95$. Guilt

and innocence interacted with type of situation, $F(1, 104) = 7.35$, (see figure 2).

Figure 2 about here

Simple main effects showed that innocent mock crime subjects scored more positively than members of any other group.

Using total score composites the police made decisions on 65% of their 40 subjects and the laboratory examiners judged 50% of their 80 subjects. The police were correct with 82% of their guilty judgements and 100% of innocent judgements. Laboratory examiners were correct with 89% and 81% of their respective guilt and innocent judgments. None of the above classification comparisons were different by chi square analyses. All of the classification methods reported in table 1 resulted in more correct than incorrect classifications by both types of examiners and in both situations using the Binomial test set at $p < .05$.

Table 1 about here

It was possible to examine the charts collected and scored by the police with the objective measurement techniques of the laboratory. Decisions by laboratory examiners were made on 62.5% of these charts.

Laboratory methods were correct on 87% of innocent decisions and 73% of guilty decisions. These detection rates were not significantly different than the rates found for the police reported above and again resulted in significantly more correct than incorrect classifications. The correlation between the scores derived by the police and by the laboratory examiners from the police charts was $r(38) = .51$. (See table 1).

The variety of possible comparisons in table one showed one significant result. Laboratory investigators made more mistakes in classifying truthful embarrassing story subjects than they did in classifying deceptive embarrassing story subjects, Fisher's exact test $p < .03$.

Table 2 shows what classification rates would be for the individual measures used by the police and laboratory investigators. Subjects with scores between ± 2 were considered to have inconclusive results whereas subjects with scores above or below those levels were classified as innocent or guilty respectively. Using the binomial expression set at the .05 level, more subjects were classified correctly than incorrectly with both types of SSR measures regardless of the examiners or the situation for which they were tested. Police investigators exceeded chance levels using their combined respiratory measure when examining mock crime subjects but not with embarrassing story subjects. Laboratory investigators successfully classified mock crime subjects but not embarrassing story subjects with thoracic respiration. Classifications were at chance levels for abdominal respiration. The blood pressure measure from the cardio arm cuff used by the police was successful for mock crime subjects but not for embarrassing story subjects.

Table 2 about here

DISCUSSION

The composite score measures for both the police and laboratory subjects differentiated between guilty and innocent subjects. The SRR measure, part of the composite in common between police and laboratory examiners, differentiated between guilty and innocent subjects. Respiration, again part of the composite, which for laboratory examiners was scored from the thoracic measure whereas for the police was derived from a blend of the abdominal and thoracic measures, differentiated between guilty and innocent subjects. The SRR results were strong enough to be reflected in accurate classifications by police and laboratory examiners in both the mock crime and embarrassing stories situations. The respiration differences translated into accurate classification results with the police and laboratory examiners but only with mock

crime tests. Heart rate, used solely by police examiners, was effective with mock crime classifications.

There was an examiner's effect with the composite scores that showed that the police in comparison to the laboratory examiners tended to score subjects more in the innocent direction. Although there was no interaction with composite scores, the SRR score results showed an interaction indicating that laboratory examiners scored embarrassing story subjects in general towards the guilty end of the continuum. These underlying results were reflected in classifications made on composite and SRR scores such that laboratory examiners made more false positive errors with embarrassing story subjects than they made with mock crime subjects.

The above examiner's effect suggests that, as laboratory testers move away from their familiar mock crime paradigm, they make more false positive errors whereas the police examiners remain consistent across different situations. It is particularly interesting that the argument presented by Iacono and Patrick

(1988) suggesting that accuracy rates could decline as the paradigms become more like real situations receives some support from the laboratory examiners but does not from the police examiners. Police work deals exclusively with real events and they would have much more experience with emotional or highly stressed suspects. From their experience they may be able to more effectively create or present the CQT test.

It is worth noting that we have little more than face validity evidence to suggest that the use of the embarrassing stories paradigm is appropriate or possibly more appropriate to study the validity of lie detection but by definition the stories deal with real events whereas the mock crime does not. In addition, Bradley, Cullen & Carle (1993) reported emotional ratings of embarrassing stories on such emotions as embarrassment, anger and anxiety and they were higher than those for the mock crime scenario.

The current police results can be compared to some field results reported by Iacono & Patrick (1988). They found 100% of guilty and 90% of innocent subjects in confession verified cases were classified correctly

by the original examiner. Blind rescoring of the charts found a 98% correct classification of guilty subjects but only a 55% correct classification of innocent subjects. These results indicate that the combination of an investigation procedure, and an informed examiner conducting the polygraph examination was very effective. The scoring of the charts alone, however, without investigative information or the personal contact and all that such contact entails yielded a result indicating that the test is biased towards the false positive error of classifying suspects in general as guilty.

Disagreement between scorers, such as that found by Iacono and Patrick (1988), has lead Furedy (1993) to question the basis of detection. How much is due to the physiological data, prior investigative information, the examiner's subjective impression or potential interactions amongst these factors? Ben-Shakhar and Furedy (1990) devoted a chapter to convince the reader that a proper analysis of the validity of the CQT would involve the discovery of how much the collection of noncontaminated physiological information

would add to interrogation procedures.

The fact that the police in our study were accurate and did not show a bias towards false positive errors in light of the above commentary becomes very important. Though they classified subjects with the same level of accuracy as the original investigators in Iacono and Patrick's (1988) report, their actual status would be somewhere in between those investigators and blind scorers. They had a description of the events on which deception might be attempted but they did not have any personal information on the suspects. Except to explain the procedures and go over the questions, there was very little interaction between the police and their suspects. In addition, there was no follow-up interview after the tests. Our results indicate that the police do not need investigative evidence or an extensive interview to achieve high levels of accuracy and avoid a bias towards finding false positive errors. It does not answer Ben-Shakhar and Furedy's (1990) question of how much more effective the addition of a polygraph test makes general interrogation procedures but it suggests that the

testing situation virtually on its own can be effective.

The discussion associated with polygraph testing has been stated in such strong terms as to be characterized as "heated debate"(p120, Dawson, 1990). Very influential authors, such as Lykken (1981) and Ben-Shakhar and Furedy (1990) have argued strongly that factors inherent as well as beyond any given Control Question test influence the outcome.

Lykken's (1981) arguments stem from his opinion that most suspects, regardless of whether they are guilty or innocent, should be more reactive to questions about a crime that they are accused of than to control questions. The evidence is mixed (eg. Kircher & Raskin, 1988) but Lykken (1981) proceeds as if he is correct and combines his opinion with some selected cases from his experience. For example, he believes that lie detection tests may be offered by prosecutors who have a weak case against the accused with the objective that if the suspect fails "then the weak case becomes suddenly much stronger" (p120, Lykken, 1981). In another example he proposes a law

entitled "Lykken's law" and applies it to a polygraph situation. The law states that when humans are faced with difficult decisions they will give greater than deserved weight to seemingly simple "objective" indicators such as the polygraph. Therefore he writes that "an accused policeman" with "a spotless record" may be considered guilty, even if innocent, because the polygraph finds him so (p69).

Ben-Shakhar and Furedy (1990) have taken one of Lykken's (1974) ideas that the lie detector is a "psychological rubber hose" for inducing confessions. To create the proper psychological set for reading their book in the preface, and on page 2, Ben-Shakhar and Furedy (1990) compare the polygraph procedures to the confession inducing function of medieval torture techniques. This kind of writing is very exciting and topical but these authors and Lykken have so freely combined imaginative social criticism with their empiricism that it is often difficult to know which statements are objectively based and which are not.

Because of their conviction that the technique cannot and does not work as a test, Lykken (1981) and

Ben-Shakhar and Furedy (1991) have created motives and reasons for the behavior of those who practice polygraph testing. Although they do not make much sense in terms of the general goals and policies of testing, these motives cover a range of possible uses that could be imagined to apply to particular cases. The examples, in the previous paragraph, that portray polygraph operators as aggressive criminal catchers who will go to the extreme of creating the appearance that a suspect is a criminal even if it is very uncertain that he is guilty has some plausibility. There may even be cases of this happening but it makes no sense to see this as anything other than isolated abuse. They give minimal consideration to the idea that a polygraph operator could be concerned with accuracy for reasons of fairness and justice.

We asked the police in this study about their results and especially the fact that their underlying scores indicated that they were biased towards judging people as innocent. We suggested that they could be letting criminals free if they make an error. In separate conversations about their results, both police

officers independently said in general that it was more important to avoid a false positive than a false negative. If someone is inaccurately found guilty of a crime, that can create a great deal of trouble for the accused person and ultimately the police officers. If they fail to find a criminal guilty, especially of a small crime, that criminal does not publicly complain and the chances are good that some other investigative evidence may turn up or that person could be picked up later on some other crime.

In general, the vigor of debate has resulted in researchers taking strong positions based on not enough research. Lykken (1981) and Ben-Shakhar and Furedy (1990) believe, derived from their rational analysis of the test, that subjects guilty or innocent will likely respond most strongly to crime relevant questions. The crime relevant questions are obvious. They are concerned with emotional events of the crime and the appearance of deception may carry severe consequences. The problem is we do not know, in the context of testing, if these authors are correct. Without really developing the theory they have put all of their faith

in explanations associated with fear of consequences and memory of emotions as the primary generator of responding.

Alternatively, if habituation of the orienting response to items in various cognitive sets was the primary factor related to responding, and emotion was a secondary factor that tended to make responses to items more resistant to habituation, then the effectiveness of the CQT is explainable. To elaborate, the police constructed the CQT in a different fashion than we did in the laboratory. The police in formulating crime relevant questions attempt to follow the "keep it simple" rule. This heuristic directs them to a single deception related issue. With the mock crime subjects, they focused on an average of 1.4 crime relevant issues with most of the questions referring to the stolen money. In comparison, our laboratory format shows we asked questions on two issues with two questions about the shooting of the victim and one about stealing money. On the embarrassing stories, they asked story relevant questions on average about 1.5 issues whereas we asked on 3 story relevant issues. There was enough

variation for the police to test what happened when they deviated from single issues. Innocent subjects scored +12 on single issue tests whereas they scored +5.6 on two issue tests ($t = 2.6$, $df = 18$)

Without falling into the trap of involved or complex explanations based on very little data, if crime relevant questions are all of the same type or on the same issue and hence are in the same category or set, whereas control questions are on a variety of issues (hence in different cognitive sets), then responses to crime relevant questions should be relatively smaller due to greater habituation. Factors such as lying, fear of consequences, vivid memories, emotional involvement, or simply personal relevance in a particular context may promote relatively greater responding by making the suspect less likely to habituate. Any or all of these factors would differentially effect the guilty suspect on the crime relevant questions. Innocent suspects may simply habituate more to crime relevant questions on a single issue because they are repeated more often than the control questions. As for the other factors with

innocent subjects, memory for the crime cannot play a role since they did not do it, also, since they did not do it the actions should be less personally relevant, lying may be associated specifically with one or more of the control questions but not with the crime relevant questions, and fear of consequences or emotionality may tend to be associated with the whole test.

Raskin (1979), building upon the work the Ben-Shakhar (1977) with relevant/irrelevant knowledge paradigms, presented a theoretical analysis using the orienting reflex. It is similar to the above but he includes the defensive reflex as a collective concept incorporating the various threatening sources of responding. He also puts a greater burden of assumptions on what the interrogator is doing in the pretest interview. The parsimonious suggestion we make is that effective tests can be constructed through ensuring that the crime relevant question remain substantially the same whereas the control (comparison) questions cover different incriminating topics. Our focus is on the idea of different comparison (control)

questions and therefore removes the burden of balancing the psychological/emotional impact between questions that Furedy (1993) suggests is necessary for "scientific control". Further because so many laboratory investigations (e.g. Raskin and Kircher, 1988) have reported success in classifying guilty subjects as guilty, it would be premature to suggest that threat value or emotional memories were necessary components to promote responding in subjects. It is possible that a sufficient condition for differential delays in habituation simply has to do with the creation of strong contextual personal relevance for crime relevant questions. This could be done in a variety of ways. For example, we have started to collect real stories written by subjects instructed to give us a very pleasant, unpleasant, or relatively neutral account of an event in their life. If negative emotions are important then subjects should be most reactive to questions on unpleasant events. If personal relevance is the major factor then subjects should react to their own story regardless of the emotional valence.

In general, experiments readily come to mind for this approach. What would be the patterns of habituation for guilty and innocent subjects through the successive presentation of the same crime relevant question? If control questions were changed in the CQT upon each presentation, would that create more false-negatives? If varying control questions is key, must they be intimidating, incriminating, embarrassing or ambiguous or is a change of topic enough to be evocative of a response? If a meta-analysis of studies that report the number of topics dealt with by crime relevant and control questions were done, would it show that the number of innocent judgments increase as the relative number of different topic control questions increase?

In sum, the present study found that both police examiners and laboratory workers were able to correctly classify subjects suspected of lying. The accuracy of classification tended to drop for laboratory investigators but not for the police when dealing with embarrassing stories. Examination of question construction revealed a difference in the number of

story or crime relevant questions asked between the police and laboratory investigators and lead to a habituation explanation for CQT accuracy.

This study was supported by a Grant from the Defense Personnel Security Research Division of the office of Naval Research of the United States.

We wish to thank Officers William Reid of the Saint John City Police Force and Steven Palmer of the Rothesay Regional Police Force for their polygraph testing and their insight into issues raised by this project.

Requests for reprints should be addressed to M. T. Bradley, Department of Psychology, University of New Brunswick, P.O. Box 5050, Saint John, New Brunswick E2L 4L5, Canada.

References

- Bradley, M.T. & Cullen, M. (1993). Polygraph lie detection on real events in a laboratory setting. Perceptual and Motor Skills, 76, 1051-1058.
- Ben-Shakhar, G. (1977). A further study of the dichotomization theory in the detection of information. Psychophysiology, 15, 408-413.
- Ben-Shakhar, G. & Furedy, J.J. (1990). Theories and Applications in the Detection of Deception. New York, Springer-Verlag.
- Dawson, M. (1990). Where does the truth lie? A review of The polygraph test: Lies, truth, and science. Psychophysiology, 27, 120-121.
- Furedy, J.J. (1993). Invited reply to Honts. Psychophysiology, 30, 319-321.
- Iacono, W.G. & Patrick, C.J. (1988). Assessing deception: Polygraph techniques. In R. Rogers (Ed.), Clinical assessment of malingering and deception (pp. 205-233). New York: Guilford.

- Kircher, J.C., & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in the laboratory setting. Journal of Applied Psychology, 73, 291-302.
- Lykken, D. T. (1981). A tremor in the blood: uses and abuses of the lie detector. New York: McGraw-Hill.
- Raskin, D. C. (1979). Orienting and defensive reflexes in the detection of deception. In H. D. Kimmel, E. H. Van Olst, and J. F. Orlebeke (Eds.), The orienting reflex in humans. Hillsdale, N. J.: Erlbaum, pp. 587-605.
- Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. American Psychologist, 40, 355-366.
- Timm, H.W. (1982). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the Guilty Knowledge Technique. Journal of Applied Psychology, 67, 391-400.

Table 1
Classification of Subjects as Guilty or Innocent Based
on Composite Scores.

Measure, examiner and actual condition		Classification					
		Correct		Incorrect		Inconclusive	
Composite Score		Story	Crime	Story	Crime	Story	Crime
Police	Guilty (40s)	5	4	1	1	4	5
	Innocent	7	8	0	0	3	2
Lab	Guilty (80s)	10	7	0	2	10	11
	Innocent	5	12	4	0	11	8
Police with lab scoring	Guilty	5	5	1	0	4	5
	Innocent	7	5	0	2	3	3

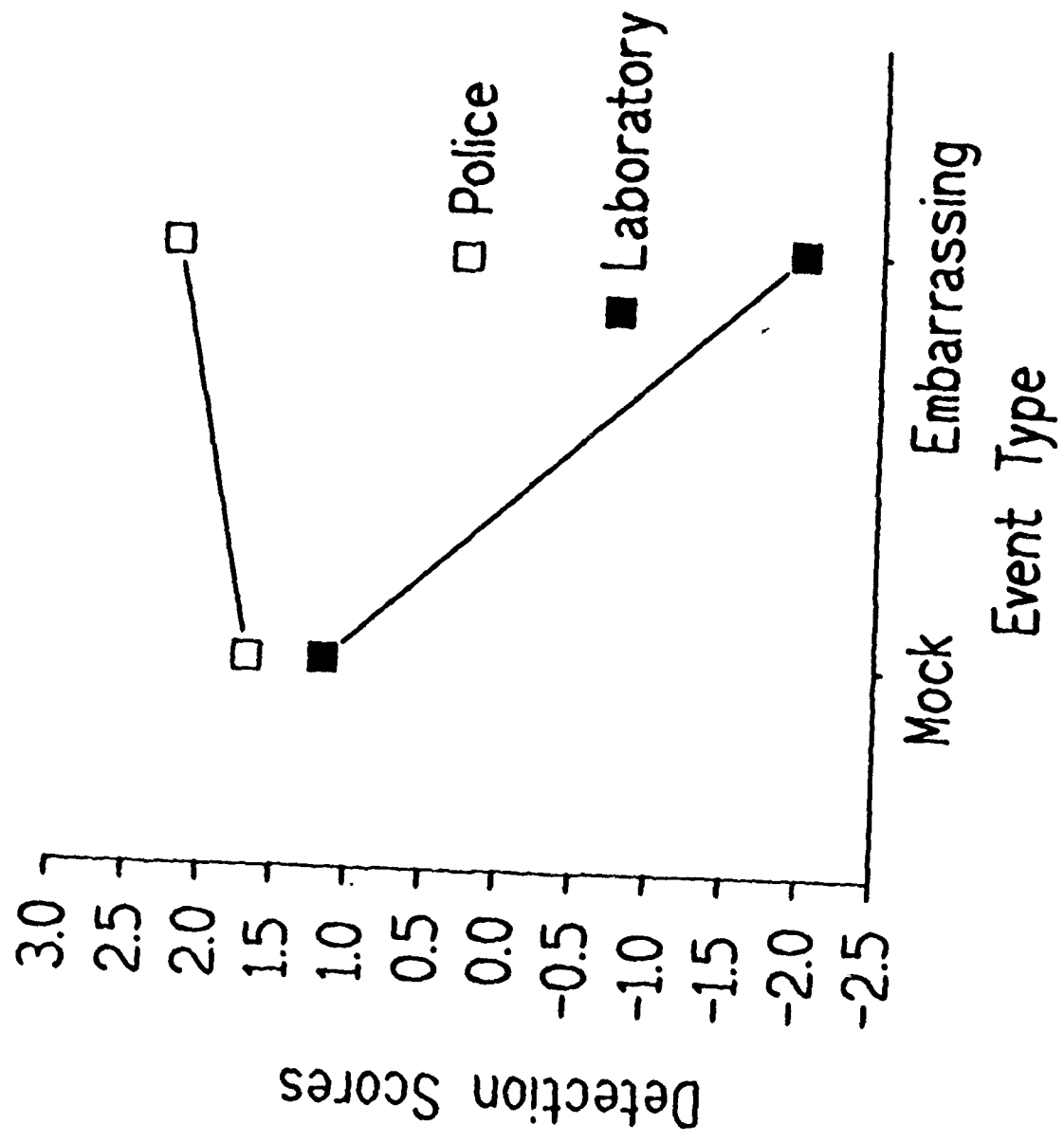
Table 2
Classification of Subjects on Individual Physiological Measures.

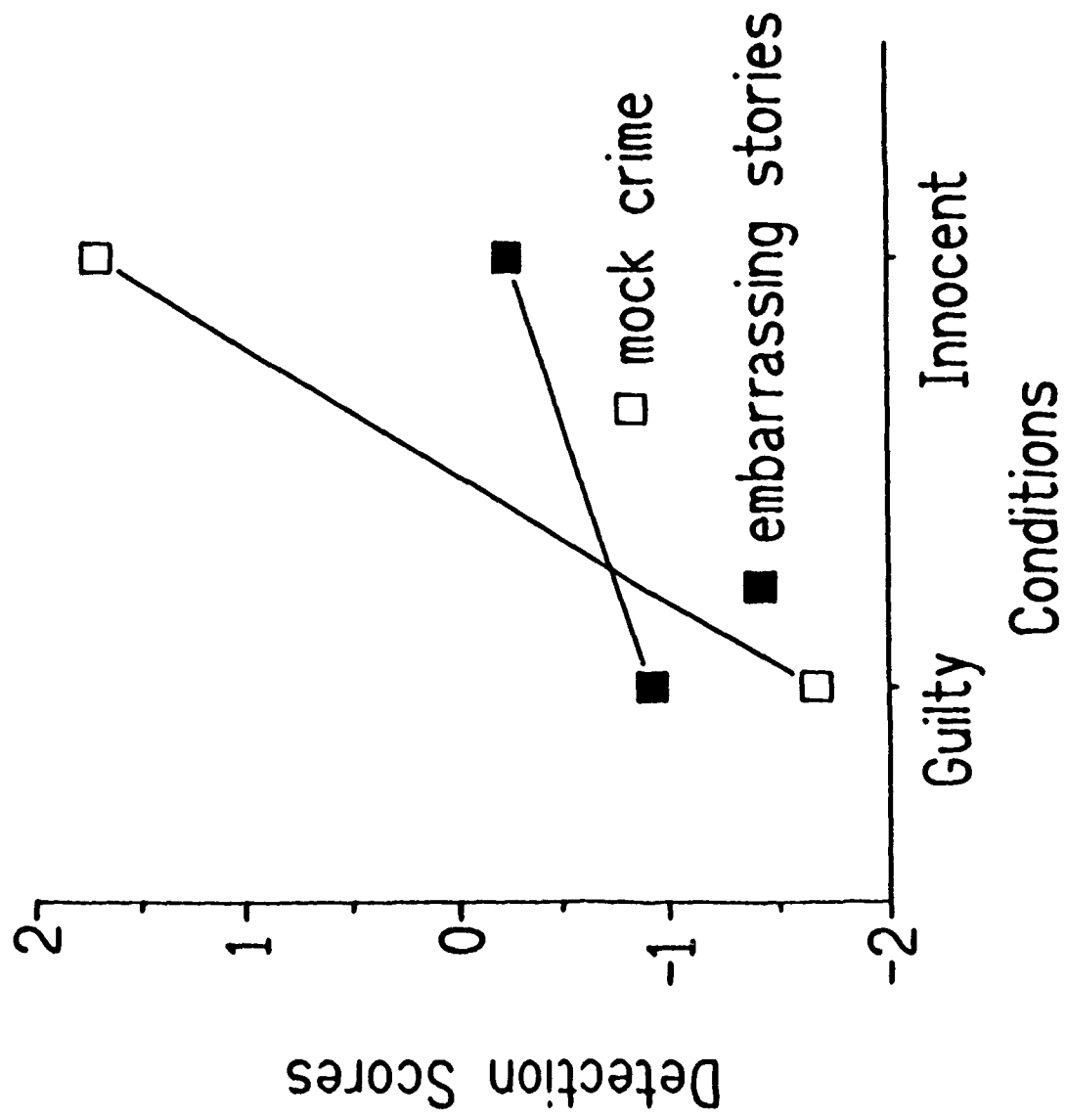
Measure, examiner and actual condition		Classification					
		Correct		Incorrect		Inconclusive	
		Story	Crime	Story	Crime	Story	Crime
SRR scores +/-1							
Police	Guilty (40s)	4	3	2	2	4	5
	Innocent	5	8	0	0	5	2
Lab	Guilty (80s)	17	10	0	3	3	7
	Innocent	3	9	4	2	13	9
SRR scores +/-3							
Police	Guilty (40s)	3	4	2	2	5	4
	Innocent	8	8	0	0	2	2
Lab	Guilty (80s)	14	8	0	4	6	8
	Innocent	4	12	6	2	10	6
Respiration							
Combo	Guilty	2	3	0	0	8	7
Police	(40s)						
	Innocent	2	3	1	0	7	7
Thor	Guilty	4	8	2	2	14	10
Lab	(80s)						
	Innocent	3	9	6	1	11	10
Abdom	Guilty	5	4	5	1	10	15
Lab	(80s)						
	Innocent	6	7	3	5	11	8
Heart rate							
Police	Guilty (39s)	5	4	2	1	3	4
	Innocent	6	7	3	0	1	3

Figure Captions

Figure 1. Interaction between event type and examiner with SRR scores.

Figure 2. Interaction between guilt condition and event type with respiration scores





Control Question Tests by Police and Laboratory
Polygraph Operators on a Mock Crime and Real Events.

M. T. Bradley, M. C. Cullen,

S. B. Carle

University of New Brunswick Saint John

Running head: POLICE AND LABORATORY POLYGRAPHERS COMPARED

Abstract

Males and females, truthful or deceptive, about a real life embarrassing story or a laboratory mock crime were examined with Control Question detection of deception tests. Exams were conducted either by a police or a laboratory trained polygraph operator. Subjects were more reactive to event relevant questions when deceptive than when truthful. Police scored subject records more towards innocence whereas laboratory investigators scored them more towards guilt. This was especially pronounced with SRR measurement on embarrassing stories. Such a result could mean that laboratory investigators when mistaken would have a tendency to classify innocent people as guilty when dealing with real events whereas the police when wrong would tend to classify the guilty as innocent.

Control question tests by police and laboratory polygraph operators on a mock crime and real events.

A number of attempts to assess criterion validity of the Control Question Test have been made in both field and laboratory studies. A recent review (Ben-Shakhar & Furedy, 1990), suggested that validity issues have not been fully resolved because of various problems particular to each area of study. A major problem in the field is that it is difficult to verify who is actually guilty or innocent by any satisfactory criteria outside of the polygraph test. Therefore, test accuracy levels cannot be determined with confidence.

This particular problem is avoided in laboratory studies because subjects can be assigned to their conditions but other problems arise. Laboratory studies are simulations of crimes. Usually, these simulations involve relatively small incentives and the population (students) participates in an exercise that may not generalize to field situations (Saxe,

Dougherty, & Cross, 1985). Laboratory examiners are typically highly trained in psychology, psychophysiological measurement and general testing, whereas field investigators are specifically trained for criminal polygraph work and have general criminal interrogation experience.

Bradley and Cullen (1993) selected one area of difference and attempted to add realism to the laboratory situation by examining events that had actually occurred to subjects. Subjects were asked to provide, from their own life, an embarrassing story that had a strong emotional impact on them. The story, which for ethical reasons, had to be non-criminal, involved events that subjects preferred no-one knew of and they would rather deny. Subjects, examined with the Control Question Test on two stories, one in which they were the principal actor and one in which they had no part, were accurately classified as deceptive in denying their own story and as truthful when denying another story.

The present study furthered explored the use of real events by comparing the results of real event

examinations with those from a mock crime situation. In addition, both police and laboratory trained examiners tested subjects.

The two police officers had been trained by the Ottawa Canadian Police College in the early 1980's. Their work since that training has been in the use of the CQT for criminal investigation. In a comparable way to the laboratory examiners, the police officers agreed to blindly examine subjects solely on the basis of knowing only the details of the mock crime or the particular embarrassing story to classify whether subjects were deceptive or truthful about their role in these events. Beyond that, the police were free to apply the CQT in the way that their training and experience dictated that they should. The laboratory examiners were restricted to a laid out protocol.

The scores of subjects examined on embarrassing stories were compared with those of subjects examined in a typical mock crime situation. This provided a direct test of conditions hopefully closer to actual field conditions as versus the enacted artificial mock crime. If considerations by Iacono and Patrick (1988)

are correct then, the accuracy of detection rates from embarrassing incidents should be somewhat less than those found in the mock crime situation.

To find out if training or experience makes a difference, results from subjects examined by police polygraph operators were compared with those tested by laboratory trained operators.

Method

Subjects

One hundred and twenty male and female introductory psychology student volunteers participated for a bonus course credit. Prior to volunteering, they were, through a consent sheet, informed of most of the experimental requirements. A sensitive issue highlighted in the form involved the fact that a limited number of people who assisted with the experiment would be able to associate their name with their embarrassing incident.

Apparatus

A Lafayette model 760-566 polygraph was used to record skin resistance responses (SRR) and respiration. Skin resistance was measured by standard Lafayette zinc chloride electrodes. After the skin had been cleaned with a cotton swab dipped in alcohol, the electrodes were attached to the medial phalanges of the first and second fingers. Respiration, both thoracic and abdominal, were measured by a standard Lafayette pneumatic chest assembly. Baseline and sensitivity levels were adjusted individually.

Procedure

Forty three male and female volunteer subjects were asked to write out in some detail a very embarrassing incident in which they were involved. The stories were read for clarity and understanding. The authors of the thirty selected stories were contacted and polygraph examination sessions were arranged. An equal number of subjects were contacted who had not written a story. They were examined on one of stories generated by the first group of subjects. Subjects who appeared truthful on the subsequent polygraph test

received \$20.00.

A second set of subjects followed instructions leading them to be guilty or innocent of a mock crime murder. Guilty subjects were asked to go into a room labelled hotel, pick up a gun from the window ledge, and shoot a mannequin wearing a red shirt three times in the chest. The mannequin was wearing a name tag with "Bob" written on it and had \$15 in his shirt pocket. Guilty subjects stole the \$15, put the money in their footwear, hid the gun in a wastebasket and left the room. They had about 10 minutes to complete their instructions and once done they went go to a room to await the return of an experimenter who arranged for a polygraph test. Subjects were told that if they appeared innocent of the crime they would receive \$20.00.

The instructions for the innocent subjects informed them that they were murder suspects and, although they had no alibi to account for their activities, they were given a chance to prove their innocence on the polygraph. These subjects were informed that they would receive \$20.00 for a judgment of innocent on the

polygraph test.

All subjects were reminded that during the polygraph examination they were, depending on the group, to deny their involvement in the mock crime or the embarrassing incident. In that way half of the subjects were deceptive and half were truthful about the events. Subjects were reminded at this point that they could receive \$20.00 for a judgment of innocent whereas they would not receive money if they were judged guilty.

The polygraph examinations, done by a laboratory examiner blind to the whether subjects were attempting deception or not, were standard for everyone in the mock crime groups. The examinations were individually constructed for each subject in the embarrassing event groups. For example, a subject was examined on vomiting over 1st class passengers in an airplane while he was running for the washroom. The Control Question Test was as follows:

- (1) Your first name is _____?
- (2) Are you going to tell me the entire truth in this examination?
- (3) Do you believe I will only ask the questions

that we have reviewed in this test?

- (4) Have you ever been arrested?
- (5) Did your embarrassing event take place on an airplane?
- (6) Have you ever cheated on your income tax?
- (7) Were you walking through the first class compartment when you vomited?
- (8) Is your last name _____?
- (9) Did you ever wish to seriously hurt a friend?
- (10) Did you make a mess on the passengers in the plane?

The standard mock crime Control Question Test was as follows:

- 1. Is your last name _____?
- 2. Regarding the crime , do you intend to answer all questions about it truthfully ?
- 3. Are you convinced that I will only ask questions on this test that we have already okayed ?
- 4. Before the age of 18, did you ever steal any money ?
- 5. Did you take the money from the body of the man?

6. While you were in school, did you ever steal anything of value ?
7. While you were in the hotel, did you shoot the man?
8. Is your first name _____?
9. Before the age of 18, did you ever physically harm anyone in any way ?
10. Did you use the gun to kill the man ?

The initial three questions and question number 8 pertained to general issues. Questions 4, 6, and 9 were control questions and questions 5, 7, and 10 were event relevant. Control questions focus on possibly incriminating issues that are not the true concern of the investigator. They are meant to be emotionally evocative, however, because they are about issues that, in this study, are related to other embarrassing situations.

The question set during the actual examination was repeated three times. After each question, approximately 20 seconds was allowed for physiological responses.

The police officers did not follow their normal

procedure involving pretest interviews as they had no investigative evidence about the suspect. They modified the mock crime CQT by having all three crime relevant questions concentrate on what they considered the single salient issue, the taking of the money. To illustrate, the following are the crime relevant questions from one test. "Concerning the case, did you take the money belonging to Bob?;, do you have Bob's money in your shoe?;, are you hiding Bob's money in your shoe?".

In a similar fashion, in general, the police focused on a single issue with embarrassing stories. The following example illustrates crime relevant questions from a story. "Regarding that story, did your mom use your condoms to embarrass you in front of your friends?;, was it your mom who embarrassed you by blowing up your hidden condoms?,, did you get embarrassed when your mom blew-up your hidden condoms?"

Data Analysis

The major analyses involved 2 x 2 x 2 MANOVAs and univariate analyses on detection scores derived from

the polygraph recordings of abdominal and thoracic respiration and skin resistance. Gender, situations (mock crime, embarrassing events), and condition (innocent or guilty) were the factors analyzed. Significance for all analyses was accepted at the .05 level.

Skin resistance responses were measured at the maximum decrease in resistance in millimetres occurring within 10 seconds of the beginning of the question. To derive a numerical score responses for control and event/mock crime relevant questions were considered in pairs; the pairs being questions 4 and 5, 6 and 7, and 9 and 10. Depending on whether the size of a response to a control question was two, three, or up to four times larger than the response to the paired event-related question, a positive one, two or three was assigned to the pair. If the response to the event related question was larger, then, depending on the relative difference a negative one, two, or three was assigned to the pair. An alternate method reported in the classification table under SRR1 involved ignoring the magnitude of the difference and the simple

assignment of a +/-1 if there was a difference.

Respiration scores were derived through the use of a contour map wheel. The wheel was used to follow the curvilinear tracings that represented inhalation and exhalation and gave distance readings in millimetres. Measures were taken for 10s of chart time following question onset. Timm (1982) found respiratory suppression associated with deception. If the response to a control question was shorter than to its paired event/mock crime relevant question a +1 was assigned, if longer then a -1 assigned, and if there was no difference the score was 0.

With three sets of questions repeated three times, for each of the measures a total of 9 judgments were made and the scores had the possibility of ranging from +9 (the maximum innocent score) to -9 (the maximum guilt score). If subjects scored greater than +2 they were classed as innocent; less than -2 resulted in a guilty classification. Scores between these numbers were judged as inconclusive. When a composite of measures was created by the police or by laboratory examiners +/-6 was used for the cut points.

Results

Four factor analyses of variance were used to examine three dependent measures. The factors were examiners (police or lab), gender, situation (mock crime or embarrassing story) and condition (guilt or innocence). The dependent measures were SRR scores, thoracic respiration scores and a composite of scores. Because of differences in scoring techniques and measures the composite score for the police consisted of the sum of SRR scores plus a blended thoracic and abdominal score plus a score derived from blood pressure measurements. The composite in the laboratory involved SRR scores, and separate scores from both thoracic and abdominal respiration.

With composite scores, there was an examiners effect ($F(1, 104) = 5.54$) such that subjects tested by police obtained scores more in the positive direction ($M = 2.60$) than subjects tested by laboratory examiners ($M = -.71$). Condition effects showed that the scores of guilty subjects ($M = -4.78$) were more negative than those of innocent subjects ($M = 5.57$), ($F(1, 104) =$

60.81). No other main effects or interactions were significant.

SRR scores differed depending on whether the police ($M = 1.98$) or laboratory personnel ($M = -.40$) conducted the tests, $F(1, 104) = 7.22$. Mock crime subjects were scored in a more positive direction ($M = 1.35$) than embarrassing event subjects ($M = -.57$). Guilty subjects scored in the negative direction ($M = -2.33$) whereas innocent subjects scored in the positive direction ($M = 3.12$), $F(1, 104) = 42.74$. Embarrassing stories and mock crimes interacted with who conducted the exam $F(1, 104) = 4.38$ (see figure 1).

Figure 1 about here

Simple main effects analysis showed that laboratory examiners scored embarrassing story subjects more negatively than they scored mock crime subjects or than the police scored either type of scenario.

Respiration scores (combined by the police) differed between guilty subjects ($M = -1.30$) and innocent subjects ($M = .75$), $F(1, 104) = 16.95$. Guilt

and innocence interacted with type of situation, $F(1, 104) = 7.35$, (see figure 2).

Figure 2 about here

Simple main effects showed that innocent mock crime subjects scored more positively than members of any other group.

Using total score composites the police made decisions on 65% of their 40 subjects and the laboratory examiners judged 50% of their 80 subjects. The police were correct with 82% of their guilty judgements and 100% of innocent judgements. Laboratory examiners were correct with 89% and 81% of their respective guilt and innocent judgments. None of the above classification comparisons were different by chi square analyses. All of the classification methods reported in table 1 resulted in more correct than incorrect classifications by both types of examiners and in both situations using the Binomial test set at $p < .05$.

Table 1 about here

It was possible to examine the charts collected and scored by the police with the objective measurement techniques of the laboratory. Decisions by laboratory examiners were made on 62.5% of these charts.

Laboratory methods were correct on 87% of innocent decisions and 73% of guilty decisions. These detection rates were not significantly different than the rates found for the police reported above and again resulted in significantly more correct than incorrect classifications. The correlation between the scores derived by the police and by the laboratory examiners from the police charts was $r(38) = .51$. (See table 1).

The variety of possible comparisons in table one showed one significant result. Laboratory investigators made more mistakes in classifying truthful embarrassing story subjects than they did in classifying deceptive embarrassing story subjects, Fisher's exact test $p < .03$.

Table 2 shows what classification rates would be for the individual measures used by the police and laboratory investigators. Subjects with scores between ± 2 were considered to have inconclusive results whereas subjects with scores above or below those levels were classified as innocent or guilty respectively. Using the binomial expression set at the .05 level, more subjects were classified correctly than incorrectly with both types of SSR measures regardless of the examiners or the situation for which they were tested. Police investigators exceeded chance levels using their combined respiratory measure when examining mock crime subjects but not with embarrassing story subjects. Laboratory investigators successfully classified mock crime subjects but not embarrassing story subjects with thoracic respiration. Classifications were at chance levels for abdominal respiration. The blood pressure measure from the cardio arm cuff used by the police was successful for mock crime subjects but not for embarrassing story subjects.

Table 2 about here

DISCUSSION

The composite score measures for both the police and laboratory subjects differentiated between guilty and innocent subjects. The SRR measure, part of the composite in common between police and laboratory examiners, differentiated between guilty and innocent subjects. Respiration, again part of the composite, which for laboratory examiners was scored from the thoracic measure whereas for the police was derived from a blend of the abdominal and thoracic measures, differentiated between guilty and innocent subjects. The SRR results were strong enough to be reflected in accurate classifications by police and laboratory examiners in both the mock crime and embarrassing stories situations. The respiration differences translated into accurate classification results with the police and laboratory examiners but only with mock

crime tests. Heart rate, used solely by police examiners, was effective with mock crime classifications.

There was an examiner's effect with the composite scores that showed that the police in comparison to the laboratory examiners tended to score subjects more in the innocent direction. Although there was no interaction with composite scores, the SRR score results showed an interaction indicating that laboratory examiners scored embarrassing story subjects in general towards the guilty end of the continuum. These underlying results were reflected in classifications made on composite and SRR scores such that laboratory examiners made more false positive errors with embarrassing story subjects than they made with mock crime subjects.

The above examiner's effect suggests that, as laboratory testers move away from their familiar mock crime paradigm, they make more false positive errors whereas the police examiners remain consistent across different situations. It is particularly interesting that the argument presented by Iacono and Patrick

(1988) suggesting that accuracy rates could decline as the paradigms become more like real situations receives some support from the laboratory examiners but does not from the police examiners. Police work deals exclusively with real events and they would have much more experience with emotional or highly stressed suspects. From their experience they may be able to more effectively create or present the CQT test.

It is worth noting that we have little more than face validity evidence to suggest that the use of the embarrassing stories paradigm is appropriate or possibly more appropriate to study the validity of lie detection but by definition the stories deal with real events whereas the mock crime does not. In addition, Bradley, Cullen & Carle (1993) reported emotional ratings of embarrassing stories on such emotions as embarrassment, anger and anxiety and they were higher than those for the mock crime scenario.

The current police results can be compared to some field results reported by Iacono & Patrick (1988). They found 100% of guilty and 90% of innocent subjects in confession verified cases were classified correctly

by the original examiner. Blind rescoring of the charts found a 98% correct classification of guilty subjects but only a 55% correct classification of innocent subjects. These results indicate that the combination of an investigation procedure, and an informed examiner conducting the polygraph examination was very effective. The scoring of the charts alone, however, without investigative information or the personal contact and all that such contact entails yielded a result indicating that the test is biased towards the false positive error of classifying suspects in general as guilty.

Disagreement between scorers, such as that found by Iacono and Patrick (1988), has lead Furedy (1993) to question the basis of detection. How much is due to the physiological data, prior investigative information, the examiner's subjective impression or potential interactions amongst these factors? Ben-Shakhar and Furedy (1990) devoted a chapter to convince the reader that a proper analysis of the validity of the CQT would involve the discovery of how much the collection of noncontaminated physiological information

would add to interrogation procedures.

The fact that the police in our study were accurate and did not show a bias towards false positive errors in light of the above commentary becomes very important. Though they classified subjects with the same level of accuracy as the original investigators in Iacono and Patrick's (1988) report, their actual status would be somewhere in between those investigators and blind scorers. They had a description of the events on which deception might be attempted but they did not have any personal information on the suspects. Except to explain the procedures and go over the questions, there was very little interaction between the police and their suspects. In addition, there was no follow-up interview after the tests. Our results indicate that the police do not need investigative evidence or an extensive interview to achieve high levels of accuracy and avoid a bias towards finding false positive errors. It does not answer Ben-Shakhar and Furedy's (1990) question of how much more effective the addition of a polygraph test makes general interrogation procedures but it suggests that the

testing situation virtually on its own can be effective.

The discussion associated with polygraph testing has been stated in such strong terms as to be characterized as "heated debate"(p120, Dawson, 1990). Very influential authors, such as Lykken (1981) and Ben-Shakhar and Furedy (1990) have argued strongly that factors inherent as well as beyond any given Control Question test influence the outcome.

Lykken's (1981) arguments stem from his opinion that most suspects, regardless of whether they are guilty or innocent, should be more reactive to questions about a crime that they are accused of than to control questions. The evidence is mixed (eg. Kircher & Raskin, 1988) but Lykken (1981) proceeds as if he is correct and combines his opinion with some selected cases from his experience. For example, he believes that lie detection tests may be offered by prosecutors who have a weak case against the accused with the objective that if the suspect fails "then the weak case becomes suddenly much stronger" (p120, Lykken, 1981). In another example he proposes a law

entitled "Lykken's law" and applies it to a polygraph situation. The law states that when humans are faced with difficult decisions they will give greater than deserved weight to seemingly simple "objective" indicators such as the polygraph. Therefore he writes that "an accused policeman" with "a spotless record" may be considered guilty, even if innocent, because the polygraph finds him so (p69).

Ben-Shakhar and Furedy (1990) have taken one of Lykken's (1974) ideas that the lie detector is a "psychological rubber hose" for inducing confessions. To create the proper psychological set for reading their book in the preface, and on page 2, Ben-Shakhar and Furedy (1990) compare the polygraph procedures to the confession inducing function of medieval torture techniques. This kind of writing is very exciting and topical but these authors and Lykken have so freely combined imaginative social criticism with their empiricism that it is often difficult to know which statements are objectively based and which are not.

Because of their conviction that the technique cannot and does not work as a test, Lykken (1981) and

Ben-Shakhar and Furedy (1991) have created motives and reasons for the behavior of those who practice polygraph testing. Although they do not make much sense in terms of the general goals and policies of testing, these motives cover a range of possible uses that could be imagined to apply to particular cases. The examples, in the previous paragraph, that portray polygraph operators as aggressive criminal catchers who will go to the extreme of creating the appearance that a suspect is a criminal even if it is very uncertain that he is guilty has some plausibility. There may even be cases of this happening but it makes no sense to see this as anything other than isolated abuse. They give minimal consideration to the idea that a polygraph operator could be concerned with accuracy for reasons of fairness and justice.

We asked the police in this study about their results and especially the fact that their underlying scores indicated that they were biased towards judging people as innocent. We suggested that they could be letting criminals free if they make an error. In separate conversations about their results, both police

officers independently said in general that it was more important to avoid a false positive than a false negative. If someone is inaccurately found guilty of a crime, that can create a great deal of trouble for the accused person and ultimately the police officers. If they fail to find a criminal guilty, especially of a small crime, that criminal does not publicly complain and the chances are good that some other investigative evidence may turn up or that person could be picked up later on some other crime.

In general, the vigor of debate has resulted in researchers taking strong positions based on not enough research. Lykken (1981) and Ben-Shakhar and Furedy (1990) believe, derived from their rational analysis of the test, that subjects guilty or innocent will likely respond most strongly to crime relevant questions. The crime relevant questions are obvious. They are concerned with emotional events of the crime and the appearance of deception may carry severe consequences. The problem is we do not know, in the context of testing, if these authors are correct. Without really developing the theory they have put all of their faith

in explanations associated with fear of consequences and memory of emotions as the primary generator of responding.

Alternatively, if habituation of the orienting response to items in various cognitive sets was the primary factor related to responding, and emotion was a secondary factor that tended to make responses to items more resistant to habituation, then the effectiveness of the CQT is explainable. To elaborate, the police constructed the CQT in a different fashion than we did in the laboratory. The police in formulating crime relevant questions attempt to follow the "keep it simple" rule. This heuristic directs them to a single deception related issue. With the mock crime subjects, they focused on an average of 1.4 crime relevant issues with most of the questions referring to the stolen money. In comparison, our laboratory format shows we asked questions on two issues with two questions about the shooting of the victim and one about stealing money. On the embarrassing stories, they asked story relevant questions on average about 1.5 issues whereas we asked on 3 story relevant issues. There was enough

variation for the police to test what happened when they deviated from single issues. Innocent subjects scored +12 on single issue tests whereas they scored +5.6 on two issue tests ($t = 2.6$, $df = 18$)

Without falling into the trap of involved or complex explanations based on very little data, if crime relevant questions are all of the same type or on the same issue and hence are in the same category or set, whereas control questions are on a variety of issues (hence in different cognitive sets), then responses to crime relevant questions should be relatively smaller due to greater habituation. Factors such as lying, fear of consequences, vivid memories, emotional involvement, or simply personal relevance in a particular context may promote relatively greater responding by making the suspect less likely to habituate. Any or all of these factors would differentially effect the guilty suspect on the crime relevant questions. Innocent suspects may simply habituate more to crime relevant questions on a single issue because they are repeated more often than the control questions. As for the other factors with

innocent subjects, memory for the crime cannot play a role since they did not do it, also, since they did not do it the actions should be less personally relevant, lying may be associated specifically with one or more of the control questions but not with the crime relevant questions, and fear of consequences or emotionality may tend to be associated with the whole test.

Raskin (1979), building upon the work the Ben-Shakhar (1977) with relevant/irrelevant knowledge paradigms, presented a theoretical analysis using the orienting reflex. It is similar to the above but he includes the defensive reflex as a collective concept incorporating the various threatening sources of responding. He also puts a greater burden of assumptions on what the interrogator is doing in the pretest interview. The parsimonious suggestion we make is that effective tests can be constructed through ensuring that the crime relevant question remain substantially the same whereas the control (comparison) questions cover different incriminating topics. Our focus is on the idea of different comparison (control)

questions and therefore removes the burden of balancing the psychological/emotional impact between questions that Furedy (1993) suggests is necessary for "scientific control". Further because so many laboratory investigations (e.g. Raskin and Kircher, 1988) have reported success in classifying guilty subjects as guilty, it would be premature to suggest that threat value or emotional memories were necessary components to promote responding in subjects. It is possible that a sufficient condition for differential delays in habituation simply has to do with the creation of strong contextual personal relevance for crime relevant questions. This could be done in a variety of ways. For example, we have started to collect real stories written by subjects instructed to give us a very pleasant, unpleasant, or relatively neutral account of an event in their life. If negative emotions are important then subjects should be most reactive to questions on unpleasant events. If personal relevance is the major factor then subjects should react to their own story regardless of the emotional valence.

In general, experiments readily come to mind for this approach. What would be the patterns of habituation for guilty and innocent subjects through the successive presentation of the same crime relevant question? If control questions were changed in the CQT upon each presentation, would that create more false-negatives? If varying control questions is key, must they be intimidating, incriminating, embarrassing or ambiguous or is a change of topic enough to be evocative of a response? If a meta-analysis of studies that report the number of topics dealt with by crime relevant and control questions were done, would it show that the number of innocent judgments increase as the relative number of different topic control questions increase?

In sum, the present study found that both police examiners and laboratory workers were able to correctly classify subjects suspected of lying. The accuracy of classification tended to drop for laboratory investigators but not for the police when dealing with embarrassing stories. Examination of question construction revealed a difference in the number of

story or crime relevant questions asked between the police and laboratory investigators and lead to a habituation explanation for CQT accuracy.

This study was supported by a Grant from the Defense Personnel Security Research Division of the office of Naval Research of the United States.

We wish to thank Officers William Reid of the Saint John City Police Force and Steven Palmer of the Rothesay Regional Police Force for their polygraph testing and their insight into issues raised by this project.

Requests for reprints should be addressed to M. T. Bradley, Department of Psychology, University of New Brunswick, P.O. Box 5050, Saint John, New Brunswick E2L 4L5, Canada.

References

- Bradley, M.T. & Cullen, M. (1993). Polygraph lie detection on real events in a laboratory setting. Perceptual and Motor Skills, 76, 1051-1058.
- Ben-Shakhar, G. (1977). A further study of the dichotomization theory in the detection of information. Psychophysiology, 15, 408-413.
- Ben-Shakhar, G. & Furedy, J.J. (1990). Theories and Applications in the Detection of Deception. New York, Springer-Verlag.
- Dawson, M. (1990). Where does the truth lie? A review of The polygraph test: Lies, truth, and science. Psychophysiology, 27, 120-121.
- Furedy, J.J. (1993). Invited reply to Honts. Psychophysiology, 30, 319-321.
- Iacono, W.G. & Patrick, C.J. (1988). Assessing deception: Polygraph techniques. In R. Rogers (Ed.), Clinical assessment of malingering and deception (pp. 205-233). New York: Guilford.

- Kircher, J.C., & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in the laboratory setting. Journal of Applied Psychology, 73, 291-302.
- Lykken, D. T. (1981). A tremor in the blood: uses and abuses of the lie detector. New York: McGraw-Hill.
- Raskin, D. C. (1979). Orienting and defensive reflexes in the detection of deception. In H. D. Kimmel, E. H. Van Olst, and J. F. Orlebeke (Eds.), The orienting reflex in humans. Hillsdale, N. J.: Erlbaum, pp. 587-605.
- Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. American Psychologist, 40, 355-366.
- Timm, H.W. (1982). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the Guilty Knowledge Technique. Journal of Applied Psychology, 67, 391-400.

Table 1
Classification of Subjects as Guilty or Innocent Based
on Composite Scores.

Measure, examiner and actual condition		Classification					
		Correct		Incorrect		Inconclusive	
Composite Score		Story	Crime	Story	Crime	Story	Crime
Police	Guilty (40s)	5	4	1	1	4	5
	Innocent	7	8	0	0	3	2
Lab	Guilty (80s)	10	7	0	2	10	11
	Innocent	5	12	4	0	11	8
Police with lab scoring	Guilty	5	5	1	0	4	5
	Innocent	7	5	0	2	3	3

Table 2
Classification of Subjects on Individual Physiological Measures.

Measure, examiner and actual condition		Classification					
		Correct		Incorrect		Inconclusive	
		Story	Crime	Story	Crime	Story	Crime
SRR scores +/-1							
	Guilty	4	3	2	2	4	5
Police	(40s)						
	Innocent	5	8	0	0	5	2
	Guilty	17	10	0	3	3	7
Lab	(80s)						
	Innocent	3	9	4	2	13	9
SRR scores +/-3							
	Guilty	3	4	2	2	5	4
Police	(40s)						
	Innocent	8	8	0	0	2	2
	Guilty	14	8	0	4	6	8
Lab	(80s)						
	Innocent	4	12	6	2	10	6
Respiration							
Combo	Guilty	2	3	0	0	8	7
Police	(40s)						
	Innocent	2	3	1	0	7	7
Thor	Guilty	4	8	2	2	14	10
Lab	(80s)						
	Innocent	3	9	6	1	11	10
Abdom	Guilty	5	4	5	1	10	15
Lab	(80s)						
	Innocent	6	7	3	5	11	8
Heart rate							
	Guilty	5	4	2	1	3	4
Police	(39s)						
	Innocent	6	7	3	0	1	3

Figure Captions

Figure 1. Interaction between event type and examiner with SRR scores.

Figure 2. Interaction between guilt condition and event type with respiration scores

